

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/109578>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

A Computational Study of Expressive Facial Dynamics in Children with Autism

Tanaya Guha, *Member, IEEE*, Zhaojun Yang, *Student Member, IEEE*,
Ruth B. Grossman, Shrikanth S. Narayanan, *Fellow, IEEE*

Abstract—Several studies have established that facial expressions of children with autism are often perceived as atypical, awkward or less engaging by typical adult observers. Despite this clear deficit in the quality of facial expression production, very little is understood about its underlying mechanisms and characteristics. This paper takes a computational approach to studying details of facial expressions of children with high functioning autism (HFA). The objective is to uncover those characteristics of facial expressions, notably distinct from those in typically developing children, and which are otherwise difficult to detect by visual inspection. We use motion capture data obtained from subjects with HFA and typically developing subjects while they produced various facial expressions. This data is analyzed to investigate how the overall and local facial dynamics of children with HFA differ from their typically developing peers. Our major observations include reduced complexity in the dynamic facial behavior of the HFA group arising primarily from the eye region.

Index Terms—Autism, perception, awkwardness, facial expressions, facial dynamics

1 INTRODUCTION

CHILDREN with Autism spectrum disorder (ASD) have significantly impaired social communication abilities, even those who have preserved language and cognitive skills and are commonly referred to as having high functioning autism (HFA). These social impairments are characterized by difficulties in using non-verbal cues [1], including difficulties in perceiving and producing facial expressions [2], [3]. It has been noted that individuals with HFA have difficulties processing dynamic facial information related to emotional state (a key component in social interaction) as compared to static stimuli [4]. They also display ambiguous expressions and more neutral/flat affect expressions compared to their TD counterparts [5]. Several studies have established that facial expressions of children with HFA are often perceived as atypical, awkward or less engaging by typical adult observers [5], [6], [7]. In fact, their expressions are perceived as awkward after just one second of interaction [7]. Even naïve observers, without any knowledge of the subjects' diagnosis, perceive children with HFA to be more awkward compared to their typically developing (TD) peers [7]. Despite this clear deficit in the quality of facial expression production, very little is understood about the underlying mechanisms and characteristics of facial expressions.

Atypicality in facial expressions of children with Autism has been studied through observations by humans [6], [7]. However, subtle movements causing differences in the facial dynamics between the HFA and the TD groups may not be easy to capture by visual inspection alone, which under-

scores the need for objective data analysis methods. The usefulness of computational behavior analysis has been noted in several recent studies, such as in analyzing atypicality in prosody [8], [9], [10], and in asynchronization of speech and gestures of children with autism [11]. Following this line of computational research to better understand behavioral phenotyping in autism, the work in this paper aims at quantifying atypicality of facial expressions in children with HFA, which is otherwise difficult to achieve by plain visual inspection. Our approach relies on direct measurement of facial movements during specific expressions, followed by analyses of those movement patterns.

To objectively understand facial expression-related atypicality in autism, we use direct motion capture (*mocap*) technique to record subtle facial movements in both children with HFA and TD subjects as they produced specific expressions. Mocap is a powerful technique to obtain detailed, precise description of gesture dynamics. It is widely used for supporting multimodal modeling in a variety of application domains including animation, human-machine interaction and sports [12], [13]. In this work on expressive facial movements, we use a mimicry paradigm to control task performance and variability, where subjects mimic a fixed set of facial expressions performed by actors in stimuli videos [14]. Movements of markers affixed to the face are recorded while the subjects mimicked facial expressions, and subsequently used for computationally examining the movement dynamics of the resulting facial expressions. The usefulness of computational approaches for analyzing such data was established in our previous work [15], [16], [17]. Our preliminary analysis indicated that facial dynamics of children with HFA differ from those of TD children in various objective ways, such as in exhibiting rougher head motion, higher variability within the HFA group, and lower bilateral symmetry [15], [16].

Our present study of facial expression production mech-

- T. Guha is with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, India.
- Z. Yang and S. S. Narayanan are with the Signal Analysis and Interpretation Lab (SAIL), University of Southern California, Los Angeles.
- R. Grossman is with the FACE Lab, Emerson College, Boston.

anisms in autism involves understanding the overall dynamics of the entire face, localized dynamics in specific facial regions, as well as the dynamical relationships between the movements across different facial regions.

The notion of complexity is fundamental to any dynamical system - mechanical or physiological, and can be understood as the rate at which new information is generated by the system. There is significant evidence of various physiological (dynamical) systems being associated with atypical and often reduced measures of complexity [18], [19], although the interpretation of complexity varies with the physiological parameters being studied and the developmental condition being investigated. Motivated by these past observations, we study the facial expression production system (a physiological dynamical system) in terms of its dynamic complexity. First, we consider the hypothesis whether expressive facial movement patterns in HFA represent a system of reduced complexity compared to the TD peers. Reduced complexity in the context of facial dynamics may be interpreted as partial loss of subtle movements, repetitive patterns in the dynamics, and highly correlated movements between facial regions. We investigate whether the TD group and the HFA group have similar or differing patterns of complexity in their overall facial dynamics using the multiple scale entropy (MSE) method, which provides a measure of system complexity [20]. Next, we hypothesize that subjects with HFA will exhibit lower complexity than the TD group for some, if not all, emotion conditions. To understand the dynamics of specific facial surface regions, we divide the face into smaller regions (eye, cheek, and mouth) following [21]. MSE-based complexity analysis is performed on each specified region separately. Through this analysis, we expect to identify those particular regions of the face that contribute to differences across HFA and control groups.

In case of reduced complexity in facial dynamics, one would expect higher predictability between facial regions i.e. the movement pattern of one region could be predicted well using the information from another. To understand such dynamical relationship between facial regions, i.e., co-variation in their movement patterns, we adopt a predictive modeling approach. Using the Granger causality model [22], we attempt to predict the dynamics of one facial region using the information from the other. In cases where behavior of one region can be predicted significantly well by another region, a causal dependency is established. We study the pairwise causal dependency between facial regions for the various facial expression conditions. Our hypothesis is that in case of reduced complexity in overall dynamics, there will be a stronger presence of causal dependency between facial regions.

The rest of the article is organized as follows: Section 2 provides details on the data collection process, Section 3 describes the computational methodologies we have used to analyze facial dynamics along with observations, followed by a discussion in Section 4 and conclusion in Section 5.

2 DATABASE

In this computational study, mocap technique is used to record facial movements of HFA and TD subjects as they

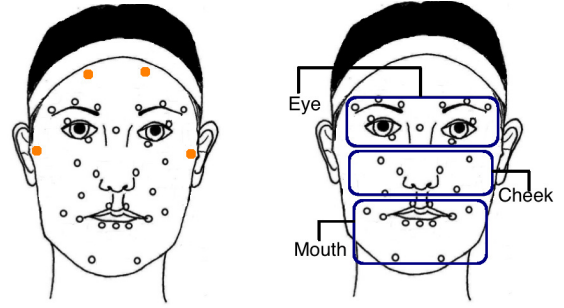


Fig. 1. Positions of the 32 facial motion capture markers affixed to each participant's face (left), Division of the 28 markers (excluding the stability markers) into eye, cheek and mouth regions (right)

produce various expressions related to emotion. To reduce task-specific variability, we adopt a mimicry paradigm, where subjects mimic a fixed set of facial expressions performed by actors in stimuli videos.

2.1 Participants

Twenty participants (2 females and 18 males) with HFA and nineteen (1 female and 18 males) TD subjects, all aged between nine and fourteen years, were recruited for this experiment at the FACE Lab at Emerson College. Diagnosis of ASD was confirmed via the Autism Diagnostic Observation Schedule, Module 3 [1] by trained administrators and confirmed by clinical impression. All participants in the ASD group demonstrated language and cognitive skills within normal limits, allowing us to describe them as having HFA. We administered the Childhood Autism Rating Scale [23] to all participants and excluded participants who scored above the threshold indicating concern for ASD from the TD cohort. Participants with learning differences (e.g. dyslexia), known genetic disorders, or other relevant diagnoses (e.g. attention deficit hyperactivity disorder) were excluded from both groups to reduce heterogeneity of the cohort. The mean ages of HFA participants (12.90 ± 3.19) and TD participants (12.67 ± 2.34) were not significantly different ($t(37) = 0.25, p = 0.80$). We assessed the subjects' IQ using the Leiter International Performance Scale [24] and receptive vocabulary with the Peabody Picture Vocabulary Test [25]. The mean IQ of HFA participants (106.35 ± 15.38) and control participants (108.74 ± 11.93) were also not significantly different ($t(37) = 0.53, p = 0.59$). Parents of all participants gave informed written consent and participants over the age of twelve provided written consent for participation in the experiment. This study was approved by the Institutional Review Board of University of Massachusetts Medical School.

2.2 Data Acquisition and Preparation

Thirty two reflective markers were affixed to the face of each participant (see Fig. 1 for marker positions). The movement of these markers was recorded by six infrared motion-capture cameras at 100 frames per second. The participants were instructed to mimic expressions in video stimuli selected from the Mind Reading corpus [14]. A study staff member was present in the room throughout the duration

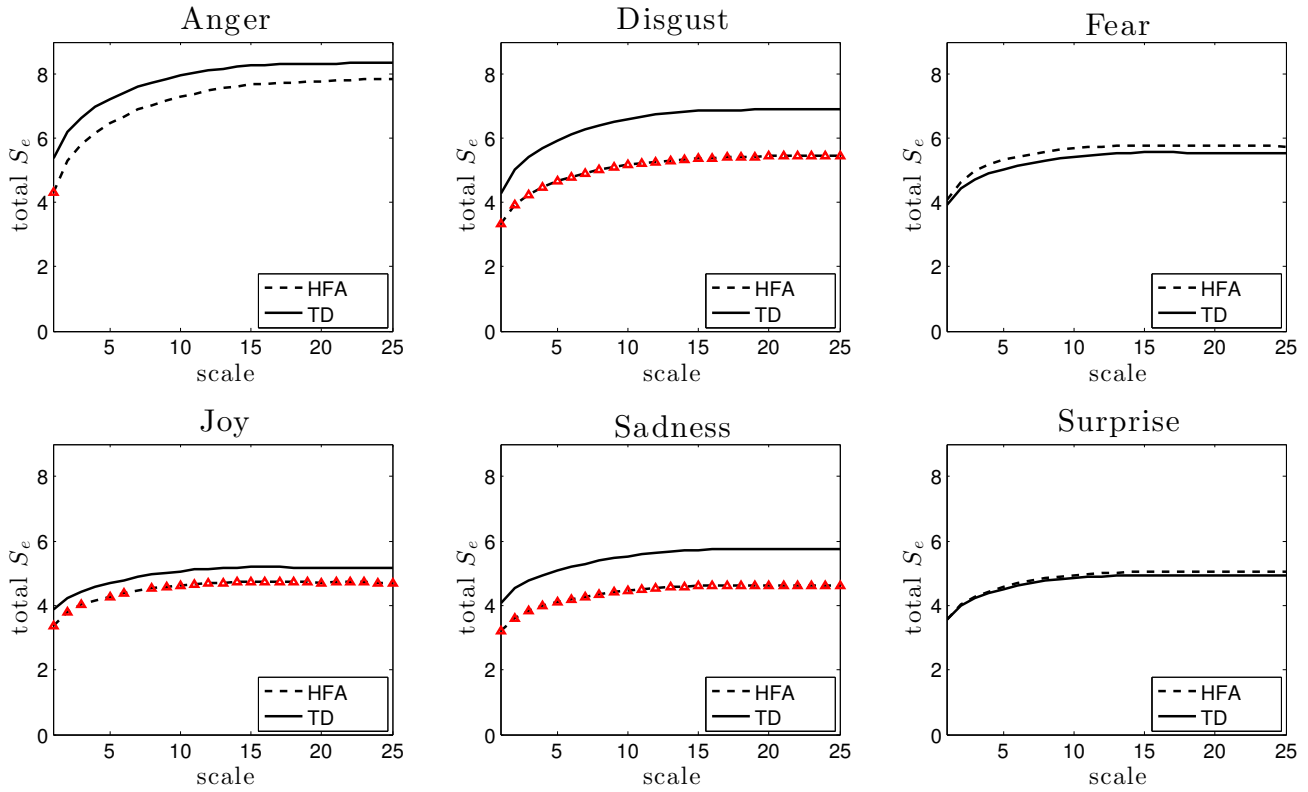


Fig. 2. Group differences in overall complexity. The red markers on the HFA plot indicate the scale factors at which HFA group has significantly lower complexity compared to the control group (error bars are omitted for clarity).

of the study to answer questions and repeat instructions if necessary.

The stimuli included two predefined, very similar sets of expressions with 18 tasks (smiling, frowning, being tearful, etc.) in each set. Each participant mimicked only one set of expressions. For ease of interpretability, these expressions are grouped into the six basic emotion categories that they are associated with - *anger*, *fear*, *disgust*, *joy*, *sadness* and *surprise*. The emotion labels for the tasks are provided in the Mind Reading corpus, and have been validated in prior research [26], [27]. Since these are not spontaneous emotions, we will refer to these categories as *facial expression condition* in this paper.

The specific arrangement of markers was based on a ninety marker, high-resolution face template [28] developed through the analysis of basic facial movement patterns using concepts of the Facial Action Coding System (FACS) [29]. We reduced the set to the 32 most critical markers delineating all central facial features, as well as overall head movement, and maintained a minimum of 4mm separation between markers to maximize recording resolution. Out of the 32 facial markers, 4 stability markers (the solid orange markers on the forehead and near the ears as shown in Fig. 1) are used to measure and later correct head motion. The positions of the remaining 28 markers are recomputed with respect to these stability markers to remove head motion. The resulting motion data are aligned, centered, and inspected manually to remove and correct artifacts. Face normalization is performed to remove subject-specific structural variability that may exist due to different facial

structures and shapes of the subjects. The processed mocap data thus accounts for pure expression-related motion, and is free from head motion and subject-specific structural variability. For more details on the preprocessing methods, refer to our previous work [16].

Each such preprocessed mocap time series, consisting of horizontal and vertical coordinates of a marker, is next converted to a distance time series $\mathcal{D} = \{d_1, d_2, \dots, d_N\}$, where d_i is the distance of the marker from its rest position at the i^{th} time instant. We use this distance time series data for further computational analyses.

3 DATA ANALYSIS AND INTERPRETATION

Our study of facial expression patterns in autism involves understanding the overall dynamics of the markers of the entire face, localized dynamics in specific facial regions, as well as the covariation of dynamics between facial regions.

3.1 Overall Facial Dynamics

In this section, we investigate whether the TD group and the HFA group have similar or differing patterns of complexity in their overall facial dynamics using the multiple scale entropy (MSE) method [20], [30]. Although the interpretation of complexity varies with the physiological parameters being studied and the developmental condition being investigated, there is significant evidence for various pathological processes being associated with atypical and often reduced measures of physiological complexity [18], [19]. We *hypothesize* that subjects with HFA will exhibit lower

complexity than the TD group for some, if not all, facial expression conditions.

The complexity of a dynamical system can be understood as the rate at which new information is produced. MSE is a measure of dynamic complexity which is often useful in the context of physiological time series. This method estimates the complexity of a time series by computing an information theoretic quantity, called the sample entropy, at multiple time scales. The sample entropy S_e at a given scale is computed as the negative logarithm of the conditional probability of two sequences within the time series being similar (in the sense of a distance metric) in an $(m + 1)$ -dimensional space, given that they are similar in the m -dimensional space.

Consider a time series $X = \{x_1, x_2, \dots, x_N\}$ from which we may form an m -dimensional vector $X_i^m = \{x_i, x_{i+1}, \dots, x_{i+m-1}\}$. Let the distance between any such two vectors be denoted as $d(X_i^m, X_j^m)$, where $i \neq j$ enforces no self-matching. The function $d(\cdot)$ in our work is the Chebyshev distance, although any distance function is applicable. Let $C^m(r)$ be the number of vector pairs for which $d(X_i^m, X_j^m) < r$ where r is a predefined threshold. Similarly, $C^{m+1}(r)$ denotes the number of cases where $d(X_i^{m+1}, X_j^{m+1}) < r$. The sample entropy is computed as:

$$S_e = -\ln \frac{C^{m+1}(r)}{C^m(r)} \quad (1)$$

For a time scale factor τ , the original time series X is first coarse-grained to obtain $Y_\tau = \{y_1, y_2, \dots, y_{N/\tau}\}$ where $y_j = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i$. For multiscale analysis, S_e is computed for multiple values of τ . In this study, we have used $m = 2$, and $r = 0.2 \times$ standard deviation of the time series. These parameter values are chosen based on previous studies that show that sample entropy has good statistical validity for these values [31].

To perform MSE analysis on our database, we use the distance time series data \mathcal{D} containing the distances of each facial marker from its rest position (see section 2.2). For each subject performing a mimicry task, we thus have a multichannel time-series, $\mathbf{D}_{task}^q \in \mathbb{R}^{N \times M}$, where q is the subject id, N is the length of the time series, and $M = 28$ facial markers. The sample entropy of \mathbf{D}_{task}^q is computed for each of the 28 channels at $\tau = 1, 2, \dots, 25$. An overall measure of complexity is obtained by adding the S_e values across all channels for each \mathbf{D}_{task}^n at each scale. If a subject performs multiple tasks under the same expression condition, the MSE values are averaged across tasks at each scale.

The overall complexity of each subject was computed in terms of the sample entropy values at 25 scale factors. To identify group difference, a two-sample t-test was performed at each scale for every expression condition at 5% significance level. In general, a system is considered more complex than the other if it produces higher values of sample entropy relative to the other consistently over increasing values of scale factor. As predicted, the results of MSE analysis shows that the HFA group has significantly reduced sample entropy patterns for three expression conditions: disgust (all 25 scales), joy (all except scales 4 and 7) and sadness (all 25 scales). Plots of group differences are presented in Fig. 2 where red markers indicate the scales

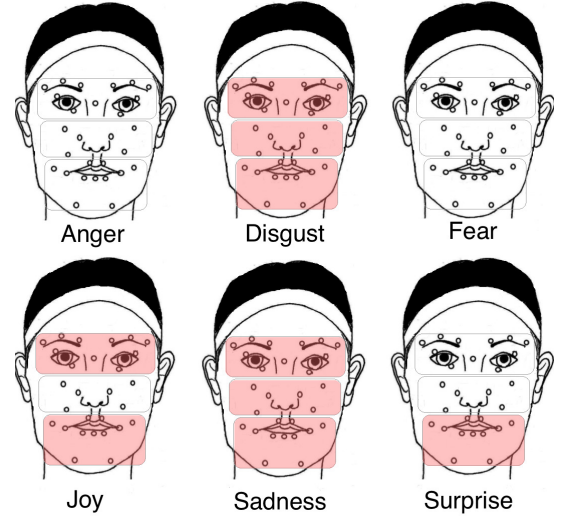


Fig. 3. Group difference in complexity for facial regions. Shaded regions indicate where the HFA group exhibit significantly lower complexity ($p \leq 0.05$) than the control group.

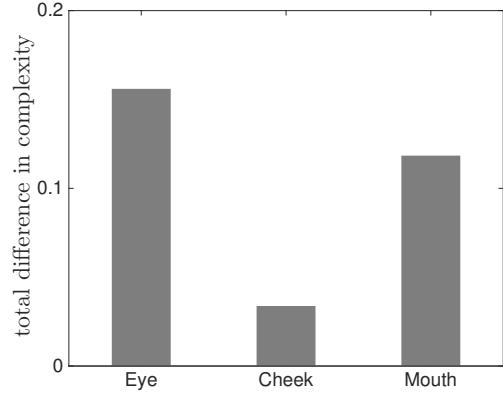


Fig. 4. Group difference in complexity for facial regions for all expression conditions

at which the group difference is significant i.e. $p \leq 0.05$. For anger, significant group difference was observed only at $\tau = 1$ which is not sufficient to infer group difference in complexity pattern. No difference is observed for fear or surprise.

3.2 Localized Facial Dynamics

It is possible that group differences in overall facial dynamics, when they exist, arise from only certain parts of the face while other facial regions exhibit normal behavior. A well established approach to studying local facial movements involves using the FACS [29], [32]. The FACS is designed to encode movement of individual facial muscles (called action units) from slight changes in facial appearance. Examples of action units are inner brow raise and lip corner pull. The facial action units thus can be considered as the low-level building block of facial expressions.

In this work, however, we do not follow the action units-based approach. Studies have reported important behavioral traits of children with autism related to facial regions - the eye avoidance hypothesis in autism [33], [34], [35] for

TABLE 1
Region of highest complexity

	HFA	TD
Anger	cheek	eye
Disgust	mouth	cheek
Fear	cheek	cheek
Joy	cheek	eye
Sadness	cheek	eye
Surprise	cheek	eye

example. Hence, we intend to analyze facial movements at the level of facial regions. We group the markers into three facial regions, eye, cheek and mouth, pertaining to the core feature areas [36] (see Fig. 1). This kind of partition has also been adopted in [21].

3.2.1 Complexity analysis

To compute the local complexity of a region, we average the MSE values across all the markers present in that region, separately at each scale factor. Two regions are considered different if their complexity is significantly different ($p \leq 0.05$) for a majority of the scale factors.

To investigate the group differences in local regions of a face, a two sample t-test was performed on the MSE results for each of the three regions (see Fig. 3). For anger and fear, no facial region was found to be significantly different between the groups. For disgust (eye: $p = 0.008$, cheek: $p = 1.77e - 5$, mouth: $p = 0.003$) and sadness (eye: $p = 7.34e - 4$, cheek: $p = 0.025$, mouth: $p = 6.61e - 4$), reduced complexity is observed for the HFA group in all three regions. For joy, the group difference comes from eye ($p = 0.014$) and mouth ($p = 0.022$) regions. Although, the overall complexity pattern of the HFA group does not differ from that of the control group for surprise, the mouth region ($p = 0.009$) still shows significantly reduced complexity. Overall, the eye region shows the highest between-group difference in complexity, followed by mouth and cheek regions (see Fig. 4).

In addition, we report the region of maximum complexity pertaining to each expression condition for the two groups (see Table 1). We observe that apart from fear, the regions of highest complexity between the two groups are always different. While the control group produces the most complex motion in the eye region in cases of anger, joy, sadness and surprise, the highest complexity in the HFA group is observed in the cheek region for these emotion-related expressions.

3.2.2 Similarity analysis

In this section, we investigate the (dis)similarity between the HFA and control groups in terms of their facial expression dynamics in the three facial region. To measure the group difference, we employ the *dynamical time warping* (DTW) method. DTW measures the similarity (or dissimilarity) between two temporal sequences by finding the best alignment between them in terms of a warping distance. It directly compares the dynamical patterns in the sequences without making any statistical assumptions. DTW has wide applications in diverse domains, such as speech recognition [37] and emotion classification [38].

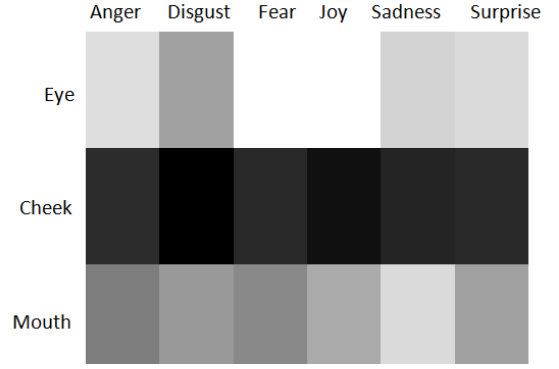


Fig. 5. Average dissimilarity in facial region dynamics between HFA and TD. Brighter color indicates larger dissimilarity.

Given two multidimensional time series $X \in R^{d \times N_x}$, $Y \in R^{d \times N_y}$, where d is the data dimensionality and N_x and N_y are the lengths of X and Y . DTW finds the best warping path by optimizing the distance between X and Y . We construct a distance matrix $\mathcal{D} \in R^{N_x \times N_y}$, where the element $d_{i,j}$ measures the distance between the i^{th} point \mathbf{x}_i in X and the j -th point \mathbf{y}_j in Y . In this work, we use ℓ_2 -norm to measure the pointwise distance: $d_{i,j} = \|\mathbf{x}_i - \mathbf{y}_j\|_2$. A warping path $W = w_1, w_2, \dots, w_K$ defines a mapping between X and Y , where the k -th element of W is defined as $w_k = (i, j)$, where $w_k(1) = i$ and $w_k(2) = j$. X can be warped to the same length of Y based on the warping path, i.e., the i^{th} point of X corresponds to the j^{th} point of Y . The optimal warping path is the one that minimizes the warping distance $d_w(W)$:

$$d_w(W) = \sum_{k=1}^K \mathcal{D}(w_k(1), w_k(2)) \quad (2)$$

This path can be found using dynamic programming with space and time complexity of $\mathcal{O}(N_x N_y)$.

We apply DTW to measure the similarity between every pair of HFA and TD subjects using the corresponding facial region-based time series data, and compute the warping distance $d_w(W)$. Fig. 5 presents the average HFA-TD similarity of facial dynamics in each region, where brighter color indicates larger warping distance, i.e., lower similarity between the groups. Consistent with our previous observation, the eye region shows a significantly larger dissimilarity than the cheek region across expression conditions ($p < 0.01$), and than the mouth region in the expression of anger, fear, joy and surprise. In addition, the cheek region exhibits the smallest dynamic difference ($p < 0.01$). The eye region-based HFA-TD difference in the expressions of joy and fear is significantly larger compared to those in other expression conditions ($p < 0.01$).

3.3 Relationship in Movement Patterns between Facial Regions

Facial expressions are produced as the results of a complex interplay and coordination between various facial regions. Such dynamic relationships may be highly non-linear, and difficult to discern with simple measures like correlation or coherence. In addition to studying the existence of the

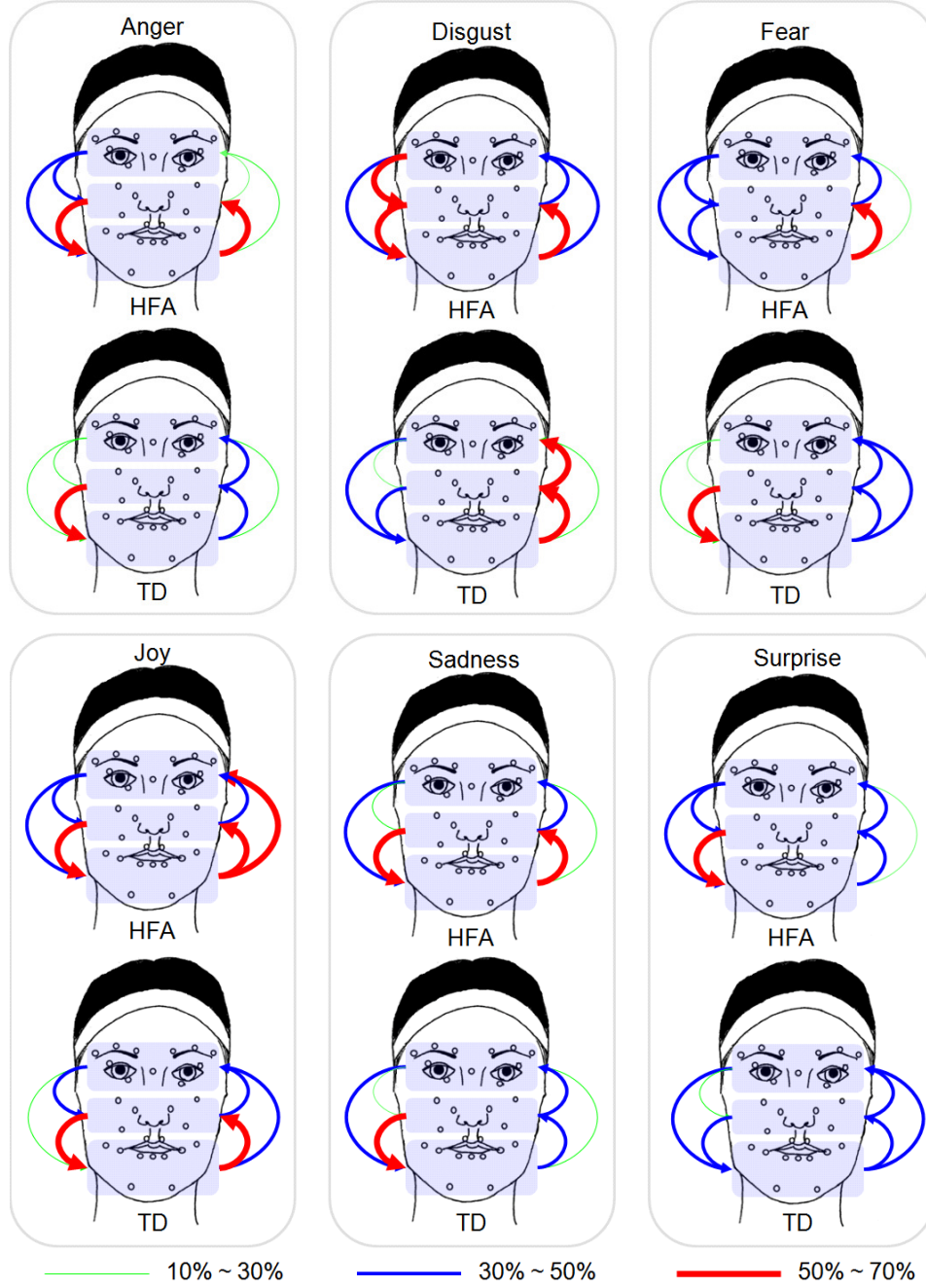


Fig. 6. Granger causality pattern between facial regions in the HFA and TD groups. The thickness and color of the arrows represent the percentage of subjects showing statistically significant causality ($p \leq 0.05$) for the corresponding region pairs.

cross-region relationship, we are interested in the interaction (directional) effects between movement patterns in different regions. To study the complex dynamic relationship between facial regions, we use the Granger causality model [22].

Granger causality is a popular statistical measure for analyzing the directional influence of one time series on another [22]. It has been widely applied to analyze various physiological and biomedical signals [39] [40] [41]. Compared to other metrics, such as correlation which is linear and symmetric, Granger causality measures a non-linear and directional relationship between variables and is more

suitable for the lag-lead system.

Given two time series X and Y , the Granger causality measure $\mathcal{F}_{X \rightarrow Y}$ defines the influence of the past information of X on improving the prediction of the current value of Y . Let's consider the following two linear regression models:

$$\begin{aligned}
 Y(t) &= \sum_{i=1}^P \alpha_i Y(t-i) + \epsilon_Y(t) \\
 Y(t) &= \sum_{i=1}^P \beta_i Y(t-i) + \sum_{i=1}^P \delta_i X(t-i) + \epsilon_{XY}(t)
 \end{aligned} \tag{3}$$

where P is the maximum number of lags for X and Y in the

models; $\{\alpha_i\}_{i=1}^P$, $\{\beta_i\}_{i=1}^P$ and $\{\delta_i\}_{i=1}^P$ are the regression coefficients, and ϵ_Y and ϵ_{XY} are the residuals of the regression models. The Granger causality measure is then defined by taking the natural logarithm of the ratio of the total variance of the two residuals ϵ_Y and ϵ_{XY} :

$$\mathcal{F}_{X \rightarrow Y} = \ln \left(\frac{\text{var}(\epsilon_Y)}{\text{var}(\epsilon_{XY})} \right) \quad (4)$$

If $\mathcal{F}_{X \rightarrow Y}$ is significantly greater than zero, then X is Granger causing Y . This significance is examined using an F -test.

For Granger causality analysis, we first determine the model order P using the Akaike information criterion (AIC). Analysis shows that 35% of the subjects attain a model order of 3, and hence we use $P = 3$ throughout our analysis. Causal dependency is examined for all possible ordered pairs of facial regions. This requires testing 6 unique ordered pairs for causality. Fig. 6 presents the Granger causality patterns between facial regions for the HFA and TD groups. The arrows indicate the direction of causal dependency originating from the region that predicts pointing to the predicted region. The thickness and the color of the arrows correspond to the percentage of subjects showing statistically significant ($p \leq 0.05$) causality between the relevant region pairs. Since the resulting percentage of subjects showing significant causality over all the pairs of facial regions ranges from 10% to 70%, we equally divide the interval [10%, 70%] into three groups to represent different strength levels between regions: low-level [10%, 30%], medium-level [30%, 50%] and high-level [50%, 70%]. As expected, we observe that subjects with HFA exhibit stronger causal dependency between regions compared to their TD peers. The dependency is particularly pronounced for the expression conditions of anger, disgust, joy and sadness.

We also observe that the causal relationship between adjacent regions is stronger than that between non-adjacent ones for both groups. For example, the coupling between cheek and mouth regions is stronger than that between eye and mouth. A strong eye-to-mouth relationship for HFA subjects is also observed for the expressions of joy.

4 DISCUSSION

The present study found reduced complexity in facial expression dynamics of subjects with HFA relative to their TD peers. Significant difference is observed for expressions related to disgust, joy and sadness. These movement differences may be the reason naïve observers perceive the facial expressions of individuals with ASD as awkward. In the context of facial expression, reduced complexity can be understood as a lack of variability, subtlety, or richness in overall facial dynamics. Our observation supports a previous study that noted more neutral/flat affect expressions in HFA group compared to their TD counterparts [5]. Our findings of reduced complexity are also consistent with past research showing atypical, often reduced, complexity measures of physiological dynamic systems in people with developmental or other conditions [18], [19].

We have also identified the facial regions with atypical complexity in HFA subjects for each expression condition. For expressions related to emotions like disgust and sadness

the reduced complexity in HFA subjects is observed across all facial regions, while for other emotions (joy and surprise) only parts of the face exhibit atypical dynamics.

On the other hand, from both local complexity and DTW-based similarity analysis, we find that the major group difference in facial dynamics comes from the eye region. Recall that the eye region exhibits the highest complexity in the TD group (see Table 1), but the HFA group shows reduced complexity in this region. The observed dynamic dissimilarity may result from this difference in complexity.

As shown in Table 1, HFA subjects exhibit highest complexity in the cheek region, which also explains the smallest dynamic difference observed in Fig. 5. These observations indicate that the cheek movements of participants with HFA are more natural and also reinforce that the perceived atypicality and awkwardness likely result from reduced complexity of movements in the eye, rather than the cheek region. One possible explanation for this behavior may come from the eye avoidance hypothesis in autism where children with autism avoid looking at the eye region of the face [33], [34], and may not be able to produce the intricate movements in the eye region, because they lack experience perceiving and processing this complex dynamic information.

The Granger causality analysis shows that HFA subjects, in general, have stronger causal dependency between facial regions compared to TD subjects. Group difference is especially prominent in the emotions of anger, disgust, joy and sadness. Recall that the HFA subjects have reduced complexity for these emotions. The strong dependency between facial regions in the HFA group suggests that subjects with HFA have smaller degrees of freedom in the underlying production mechanism, which in turn, supports our observation of reduced complexity in overall facial movements.

5 CONCLUSION

Several autism research studies have emphasized the importance of understanding vocal [9], [11] and facial behavioral expressions [7], since they have such significant impact on how children with HFA are perceived by naïve observers. In this study, we analyzed mimicked facial expressions in children with HFA using computational techniques. We observe that facial expressions of children with HFA have reduced complexity when compared to those of TD subjects. This is further emphasized by the observation of higher causal dependency between various facial regions in HFA across various expression conditions, suggesting lower degrees of freedom in the underlying mechanism. The group differences in expression dynamics are more prominent for expressions of anger, disgust, joy and sadness, and mainly arise from the eye region.

Since our observations are made based on analyzing the mocap recordings of only 20 children with autism, we cannot generalize the findings to the entire autism spectrum population, particularly due to the inherent heterogeneity in the condition. Our current experimental setup also can not identify whether the observed differences between the HFA and TD subjects resulted from an underlying group difference in perceiving the visual stimuli vs. a pure production difference. Our findings also do not speak to the

spontaneous production of expressions during naturalistic social interactions - since the expressions we analyzed are based on a mimicry paradigm. Nevertheless, within the scope of this study, our analyses suggest that children with HFA lack richness and variability in their facial expression patterns, indicating a production mechanism that allows less degrees of freedom of movement between regions.

ACKNOWLEDGMENTS

This work is supported by National Science Foundation (NSF) and National Institutes of Health (NIH). The authors would also like to thank the anonymous reviewers whose insightful comments and suggestions have enriched this article.

REFERENCES

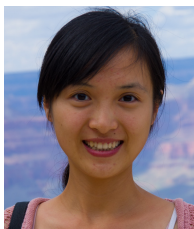
- [1] C. Lord, M. Rutter, P. C. DiLavore, S. Risi, K. Gotham, and S. L. Bishop, *Autism diagnostic observation schedule: ADOS-2*. Western Psychological Services Los Angeles, CA, 2012.
- [2] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," *Archives of general psychiatry*, vol. 59(9), pp. 809–816, 2002.
- [3] R. B. Grossman, L. R. Edelson, and H. Tager-Flusberg, "Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 3, pp. 1035–1044, 2013.
- [4] K. A. Pelphrey, J. P. Morris, G. McCarthy, and K. S. LaBar, "Perception of dynamic changes in facial affect and identity in autism," *Social cognitive and affective neuroscience*, 2007.
- [5] N. Yirmiya, C. Kasari, M. Sigman, and P. Mundy, "Facial expressions of affect in autistic, mentally retarded and normal children," *Journal of Child Psychology and Psychiatry*, vol. 30, no. 5, pp. 725–735, 1989.
- [6] S. Stagg, R. Slavny, C. Hand, A. Cardoso, and P. Smith, "Does facial expressivity count? how typically developing children respond initially to children with autism," *Autism*, pp. 704–711, 2013.
- [7] R. B. Grossman, "Judgments of social awkwardness from brief exposure to children with and without high-functioning autism," *Autism*, vol. 19, no. 5, pp. 580–587, 2015.
- [8] D. Bone, M. P. Black, A. Ramakrishna, R. Grossman, and S. S. Narayanan, "Acoustic-prosodic correlates of awkward prosody in story retellings from adolescents with autism," in *Interspeech*, 2015, pp. 1616–1620.
- [9] D. Bone, M. P. Black, C.-C. Lee, M. E. Williams, P. Levitt, S. Lee, and S. Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist," in *Interspeech*, 2012, pp. 1043–1046.
- [10] J. J. Diehl and R. Paul, "Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders," *Research in autism spectrum disorders*, vol. 6, no. 1, pp. 123–134, 2012.
- [11] A. de Marchena and I.-M. Eigsti, "Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency," *Autism research*, vol. 3, no. 6, pp. 311–322, 2010.
- [12] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P. F. Driessen, "Gesture-based affective computing on motion capture data," in *Affective computing and intelligent interaction*. Springer, 2005, pp. 1–7.
- [13] A. Metallinou, C.-C. Lee, C. Busso, S. Carnicke, and S. Narayanan, "The usc creativeit database: A multimodal database of theatrical improvisation," *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality 18 May 2010*, p. 55, 2010.
- [14] S. Baron-Cohen, *Mind reading I: the interactive guide to emotions*. Jessica Kingsley Publishers, 2003.
- [15] T. Guha, Z. Yang, A. Ramakrishna, R. B. Grossman, D. Hedley, S. Lee, and S. S. Narayanan, "On quantifying facial expression-related atypicality of children with autism spectrum disorder," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 803–807.
- [16] A. Metallinou, R. B. Grossman, and S. Narayanan, "Quantifying atypicality in affective facial expressions of children with autism spectrum disorders," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.
- [17] R. Grossman, T. Guha, Z. Yang, D. Hedley, and S. S. Narayanan, "Missing the mark: Dynamic differences in facial expressions of children with hfa," in *Int. Meeting for Autism Research (IMFAR)*, 2015.
- [18] A. Catarino, O. Churches, S. Baron-Cohen, A. Andrade, and H. Ring, "Atypical eeg complexity in autism spectrum conditions: a multiscale entropy analysis," *Clinical Neurophysiology*, vol. 122, no. 12, pp. 2375–2383, 2011.
- [19] A. L. Goldberger, C.-K. Peng, L. A. Lipsitz, D. E. Vaillancourt, and K. M. Newell, "What is physiologic complexity and how does it change with aging and disease? authors' reply," *Neurobiology of aging*, vol. 23, no. 1, pp. 27–29, 2002.
- [20] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical review letters*, vol. 89, no. 6, p. 068102, 2002.
- [21] C. Busso and S. S. Narayanan, "Interrelation between speech and facial gestures in emotional utterances: a single subject study," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 8, pp. 2331–2347, 2007.
- [22] C. W. Granger, B.-N. Huangb, and C.-W. Yang, "A bivariate causality between stock prices and exchange rates: evidence from recent asianflu," *The Quarterly Review of Economics and Finance*, vol. 40, no. 3, pp. 337–354, 2000.
- [23] E. Schloper, R. Reichler, and B. Renner, "The childhood autism rating scale western psychological services los angeles," 1988.
- [24] G. H. Roid and L. J. Miller, "Leiter-r performance scale, revised," *Wood Dale, IL: Stoelting Co*, 1997.
- [25] L. M. Dunn and L. M. Dunn, *PPVT-III: Peabody picture vocabulary test*. American Guidance Service Circle Pines, MN, 1997.
- [26] O. Golan and S. Baron-Cohen, "Systemizing empathy: Teaching adults with asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia," *Development and psychopathology*, vol. 18, no. 2, p. 591, 2006.
- [27] S. Baron-Cohen, J. Hill, and S. Wheelwright, "Mind reading made easy," *Cambridge Medicine*, 2002.
- [28] J. Hauck, D. J., "House of moves high resolution facial marker-set (92 markers)," *Los Angeles, Digital Concepts Group, inc. House of Moves*, 2007.
- [29] P. Ekman and W. V. Friesen, *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press, 1978.
- [30] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of biological signals," *Physical review E*, vol. 71, no. 2, p. 021906, 2005.
- [31] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 283, no. 3, pp. R789–R797, 2002.
- [32] Y.-l. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 2, pp. 97–115, 2001.
- [33] J. W. Tanaka and A. Sung, "The eye avoidance hypothesis of autism face processing," *Journal of autism and developmental disorders*, pp. 1–15, 2013.
- [34] D. Kliemann, I. Dziobek, A. Hatri, J. Baudewig, and H. R. Heekeren, "The role of the amygdala in atypical gaze on emotional faces in autism spectrum disorders," *The Journal of Neuroscience*, vol. 32, no. 28, pp. 9469–9476, 2012.
- [35] T. L. Hutchins and A. Brien, "Conversational topic moderates social attention in autism spectrum disorder: Talking about emotions is like driving in a snowstorm," *Research in Autism Spectrum Disorders*, vol. 26, pp. 99–110, 2016.
- [36] K. A. Pelphrey, N. J. Sasson, J. S. Reznick, G. Paul, B. D. Goldman, and J. Piven, "Visual scanning of faces in autism," *Journal of autism and developmental disorders*, vol. 32, no. 4, pp. 249–261, 2002.
- [37] C. Myers, L. R. Rabiner, and A. E. Rosenberg, "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 6, pp. 623–635, 1980.
- [38] Y. Kim and E. M. Provost, "Emotion classification via utterance-level dynamics: A pattern-based approach to characterizing affective expressions," in *Acoustics, Speech and Signal Processing, 2013 IEEE International Conference on*, 2013, pp. 3677–3681.

- [39] A. B. Barrett, M. Murphy, M.-A. Bruno, Q. Noirhomme, M. Boly, S. Laureys, and A. K. Seth, "Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia," *PloS one*, vol. 7, no. 1, p. e29072, 2012.
- [40] X. Miao, X. Wu, R. Li, K. Chen, and L. Yao, "Altered connectivity pattern of hubs in default-mode network with alzheimers disease: an granger causality modeling approach," *PloS one*, vol. 6, no. 10, p. e25546, 2011.
- [41] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by granger causality," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9849–9854, 2004.



Tanaya Guha is an Assistant Professor in the department of Electrical Engineering at the Indian Institute of Technology (IIT) Kanpur. Prior to joining IIT, she was a postdoctoral fellow at the Signal Analysis and Interpretation Lab (SAIL), University of Southern California (USC) from 2013 to 2015. She has received her PhD in Electrical and Computer Engineering from the University of British Columbia (UBC), Vancouver in 2013. She was a recipient of Mensa Canada Woodhams memorial scholarship, Google Anita

Borg memorial (finalist) scholarship and Amazon Grace Hopper celebration scholarship. Her current research interests include human-centered signal processing including human emotion and behavior analysis, and multimodal signal processing.



Zhaojun Yang is a PhD candidate in Electrical Engineering at the University of Southern California (USC). She received her B.E. Degree in Electrical Engineering from University of Science and Technology of China (USTC) 2009 and M.Phil. Degree in Systems Engineering and Engineering Management from Chinese University of Hong Kong (CUHK) 2011. Her research interests include multimodal emotion recognition and analysis, interaction modeling and spoken dialog system. She was awarded with the USC

Annenberg Fellowship (2011-2015). Her work (with S. S. Narayanan) has won the Best Student Paper Award at ICASSP 2016.



Ruth B Grossman is an associate professor at Emerson College where she teaches graduate and undergraduate classes in the department of Communication Sciences and Disorders. Her research is focused on various aspects of face-to-face communication in children with Autism Spectrum Disorders (ASD). She is specifically interested in how children with ASD integrate and produce verbal and nonverbal information, such as facial expressions and prosody. Her work has been published in several refereed

journals, including Journal of Child Psychology and Psychiatry, Journal of Speech, Language, and Hearing Research, Journal of Nonverbal Behavior, Autism, and Autism Research. More about her can be found in facelab.emerson.edu.



Shrikanth (Shri) Narayanan (StM'88 - M'95 - SM'02 - F'09) is Andrew J. Viterbi Professor of Engineering at the University of Southern California (USC), and holds appointments as Professor of Electrical Engineering, Computer Science, Linguistics, Psychology, Neuroscience and Pediatrics and as the founding director of the Ming Hsieh Institute. Prior to USC he was with AT&T Bell Labs and AT&T Research from 1995-2000. At USC he directs the Signal Analysis and Interpretation Laboratory (SAIL). His

research focuses on human-centered signal and information processing and systems modeling with an interdisciplinary emphasis on speech, audio, language, multimodal and biomedical problems and applications with direct societal relevance. [<http://sail.usc.edu>].

Prof. Narayanan is a Fellow of the Acoustical Society of America and the American Association for the Advancement of Science (AAAS) and a member of Tau Beta Pi, Phi Kappa Phi, and Eta Kappa Nu. He is the Editor in Chief for IEEE Journal of Selected Topics in Signal Processing, an Editor for the Computer Speech and Language Journal and an Associate Editor for the IEEE Transactions on Affective Computing, APSIPA Transactions on Signal and Information Processing and the Journal of the Acoustical Society of America. He was also previously an Associate Editor of the IEEE Transactions of Speech and Audio Processing (2000-2004), IEEE Signal Processing Magazine (2005-2008), IEEE Transactions on Multimedia (2008-2011) and the IEEE Transactions on Signal and Information Processing over Networks (2014-2015). He is a recipient of a number of honors including Best Transactions Paper awards from the IEEE Signal Processing Society in 2005 (with A. Potamianos) and in 2009 (with C. M. Lee) and selection as an IEEE Signal Processing Society Distinguished Lecturer for 2010-2011 and ISCA Distinguished Lecturer for 2015-2016. Papers co-authored with his students have won awards including the 2014 Ten-year Technical Impact Award from ACM ICMI and at Interspeech, 2016 ICASSP, 2015 Nativeness Detection Challenge, 2014 Cognitive Load Challenge, 2013 Social Signal Challenge, Interspeech 2012 Speaker Trait Challenge, Interspeech 2011 Speaker State Challenge, InterSpeech 2013 and 2010, InterSpeech 2009-Emotion Challenge, IEEE DCOSS 2009, IEEE MMSP 2007, IEEE MMSP 2006, ICASSP 2005 and ICSLP 2002. He has published over 700 papers and has been granted seventeen U.S. patents.